Color Rendering: A Tale of Two Metrics

Mark S. Rea,* Jean P. Freyssinier-Nova

Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY 12180

Received 2 June 2007; revised 24 August 2007; accepted 5 September 2007

Abstract: It was the best measure of color rendering, it was the worst measure of color rendering. Color rendering index (CRI) is the most common metric used by the lighting industry to represent the color rendering properties of electric light sources. CRI was intended to characterize how ''true'' or ''natural'' objects appeared when illuminated by a light source, but was never intended to, for example, represent how well object colors could be differentiated under a light source, another important aspect of color rendering. Data presented here demonstrate that CRI in conjunction with another measure of color rendering, gamut area index (GAI), is useful at predicting subjective judgments of how ''natural'' objects appear as well as how ''vivid'' objects appear, and how well one can discriminate between subtle differences in hue. Neither measure by itself, however, is sufficient for meeting all of the expectations of a light source for providing good color rendering under all viewing conditions. It remains for future research to determine if just two metrics are sufficient to assure good color rendering from a light source and whether these two metrics (CRI and GAI) are the best for such purpose. In the meantime, CRI and GAI should be used jointly in recommendations as practical, useful, and mutually reinforcing measures of color rendering. The data presented here also demonstrate that total irradiance is important for good color rendering. © 2008 Wiley Periodicals, Inc. Col Res Appl, 33, 192–202, 2008; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/col.20399

Key words: color rendering; index; gamut area; full spectrum; fluorescent; light emitting diodes; phosphor; correlated color temperature; cool; warm; white; naturalness; vividness; discrimination; Farnsworth

INTRODUCTION

P.J. Bouma in 1947 described the aspects of daylight that make it an ideal source of illumination[1]: ''It (daylight) displays (1) a great variety of colors, (2) makes it easy to distinguish slight shades of color, and (3) the colors of objects around us obviously look natural.'' His description is, in fact, an articulation of the important features of color rendering for any light source. In general, a source with good color rendering properties, daylight or electric, should reveal a full range of colors (i.e., hue, lightness, and saturation), should enable good color discrimination between objects of similar spectral reflectance (e.g., purple from maroon), and should not distort colors (e.g., should not overenhance the red appearance of hamburger in a meat case).

In the context of Bouma's early description, color rendering should be evaluated in several ways because there are several different aspects of color rendering. More strongly stated, color rendering should not be measured by any single metric. Nevertheless, we presently have only one recognized measure of color rendering in the lighting industry, color rendering index (CRI),* developed in the early 1960s through a collaborative effort among interested scientists and manufacturers.[1,3] These thought leaders developed CRI as a measure of how ''true'' objects were rendered by electric light sources.[1,3] They assumed that daylight and incandescent light should be reference illuminants in a system of color rendering because these were the most familiar sources (at least for those people in the United States) and should, thereby, render objects ''truly'' or ''naturally.'' The general CRI of all electric light sources was defined in terms of the net shift in color space of eight standard reflectances relative to the coordinates of these standard reflectances under the reference light source of the same correlated color temperature (CCT). The general CRI scale was normalized to 100 for reference illuminants and set at 50 for a warm white halophosphor fluorescent lamp.

*Correspondence to: Mark S. Rea (e-mail: ream@rpi.edu).

Contract grant sponsor: US Environmental Protection Agency; Peter Banwell served as project manager.

*For the purposes of this report, CRI is used synonymously with general CRI.[2]

FIG. 1. Farnsworth-Munsell test trial under daylight (left) and a view into the apparatus used during the study (right).

Their thinking was that because daylight and incandescent light were the most familiar light sources, the larger the shift within the color space of the standard reflectances under the test light source, relative to the chromaticity under the reference source, the less ''true'' or ''natural'' objects would appear under that source. Importantly too, these thought leaders recognized that among the limits of CRI, *any* single measure was incomplete in describing the color rendering properties of a light source. Judd, one of the authors of CRI, developed a ''flattery index'' in response to the inherent limits of CRI to characterize color rendering completely.[1,3] Judd believed that CRI could characterize how ''true'' or ''natural'' objects appeared, but could not characterize how ''flattering'' or ''vivid'' objects might appear, particularly the redness and saturation of human skin.[4]

Another important aspect of color rendering identified by Bouma, but only sporadically considered,[5–9] is the ability of a light source to reveal subtle differences in spectral reflectance (e.g., purples from maroons). Thornton promoted the concept of gamut area as a measure of ''color discrimination.'' This concept has not been embraced by the lighting industry[10] but has been the foundation of the display industry because it is important for the accurate reproduction of saturated colors. Interestingly, gamut area is also the basis for modern triphosphor fluorescent lamps developed first by Thornton when he was at Westinghouse.[11–13]

With the rapid development of solid state lighting technology, there is now a recognized need to refine, enhance, or expand CRI to satisfactorily encompass all the aspects of color rendering articulated by Bouma in 1947.[10,14,15] Based upon original empirical data presented here, a simple metric using the gamut area of the eight CIE standard reflectance samples, gamut area index (GAI, defined in the Appendix), is proposed as a practical complementary metric to the well-established CRI in characterizing the color rendering properties of electric light sources. These data demonstrate that light sources meeting *both* a CRI criterion of 80 and a GAI between 80 and 100 consistently ensure objects will appear ''natural'' and ''vivid'' and that subtle differences in hue will be perceived. It is also demonstrated here that good color rendering can only be achieved under sufficient levels of irradiance; good color rendering is not possible under low light levels.

METHODS

Three experiments were performed. Each experiment examined the color rendering properties of different phosphor-based light sources. The first experiment used only warm (low) CCTs (<4000 K), the second used only cool (high) CCTs (>5000 K), and the third used both warm and cool CCTs.

Each experiment incorporated two types of tasks: a color discrimination task and a paired-comparison task. During the first task, observers performed the Farnsworth–Munsell 100-hue test[16,17] under four different light sources and under two light levels (Fig. 1, right panel). During the second task, observers were presented with a collage of pictures showing two species of birds, blue jays and red cardinals. Observers had to compare the ''naturalness'' and the ''vividness'' separately of the blues and the reds under the different lighting conditions. Observers also made an assessment of the overall ''naturalness'' and ''vividness'' of the collage.

Subjective evaluations of color rendering can be difficult to analyze if the observers' evaluation criteria are not clearly defined (e.g., vague descriptors such as ''preference''), or if observers can evaluate different features of the same display (e.g., a multichromatic scene where some observers focus in just one color and other observers give an overall rating). To minimize these problems, the present study aimed to (1) direct the observers' attention to specific characteristics of the pictures in the collage, such as the red or blue colors, and to (2) use specific subjective evaluation criteria, such as ''naturalness'' and ''vividness.'' Although ''naturalness'' evaluations were open to the recollection and interpretation of the observers, this instruction was implemented as a reflection of the subjective criterion assumed important by the developers of CRI and was designed to help ensure a higher degree of consistency among observers in their evaluations than might otherwise have occurred with an ambiguous criterion like ''preference.''

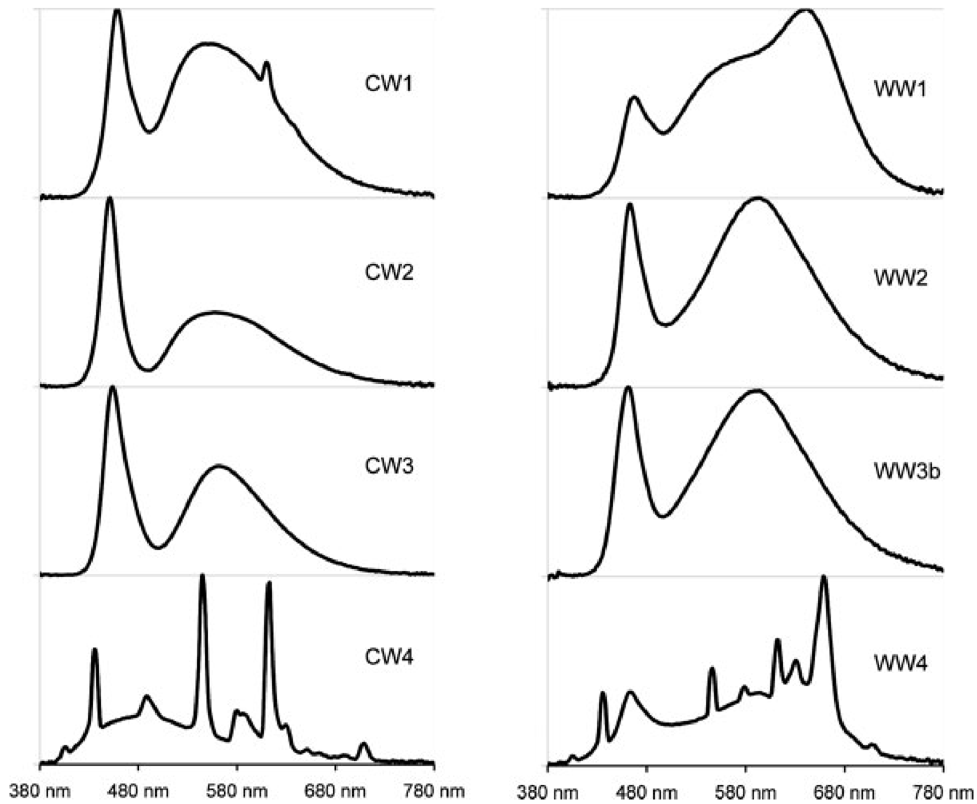

FIG. 2. Normalized spectral power distributions of the illuminants used in this study at 50 fc. Color characteristics or each illuminant are summarized in Table I.

Apparatus

A matte-white viewing cube, 2-ft on a side (Fig. 1, right panel), provided illumination from one of eight different spectral power distributions (SPDs; Fig. 2). These SPDs were produced from five commercially available phosphor-based white LEDs (Luxeon I 3000 and 5000 K; Nichia Jupiter 3300 and 6500 K; Seoul Semiconductor P3 6500 K) and two different linear fluorescent lamps (F40T12, 2900 K, Promolux Platino; and F32T8, 5000 K, Lumichrome). Two SPDs used in the experiments were obtained by mixing light from two sources.

CCT, CRI, GAI, and full spectrum color index (FSCI)[10,18] were used to characterize the color characteristics of the eight light sources (Table I). Entries in this table were obtained from actual SPDs reflected from the interior walls of the cube as determined by a calibrated spectroradiometer (PR-705; Photo Research, Chatsworth, CA).

For testing purposes, the light sources were grouped in terms of CCT. The CCTs of the sources in the first experiment (denoted as WW1 to WW4) ranged from 2900 to 3800 K (''warm white''), whereas the CCTs of the sources in the second experiment (denoted as CW1 to CW4) ranged from 5000 to 6700 K (''cool white''). In the third experiment, WW1, WW2, CW1, and CW2 were used.

Electrical dimming was used to modulate the light output of the LEDs; mechanical baffles or neutral density filters were used with the fluorescent lamps. Table I shows the resulting differences in spectral characteristics due to dimming.

The Farnsworth–Munsell 100-Hue Test consists of 85 color discs arranged in four series and is used widely as a

TABLE I. Color characteristics of the eight white illuminants used in this study.

Light source	CCT (K)	CRI	GAI	FSCI
WW1–5 fc	3157	95	52	66
WW1–50 fc	3174	95	55	67
WW2–5 fc	3399	80	64	68
WW2–50 fc	3443	80	64	68
WW3b–5 fc[a]	3707	81	73	72
WW3b–50 fc[a]	3732	81	74	72
WW4–5 fc[b]	3279	92	94	79
WW4–50 fc[b]	3261	91	93	77
CW1–5 fc	5069	75	65	60
CW1–50 fc	5137	75	66	61
CW2–5 fc	6502	78	95	73
CW2–50 fc	6682	78	97	74
CW3–5 fc	6400	72	81	63
CW3–50 fc	6126	71	81	64
CW4–5 fc	5239	94	90	72
CW4–50 fc	5854	94	99	74

The color metrics were derived from spectral power measurements taken from the walls of the experimental booth at both 5 and 50 fc.

[a] WW3b was the result of mixing 85% of WW2 and 15% of CW2.

[b] WW4 was the result of mixing 52% of WW2 and 48% of a F40T12 3000 K Promolux Platino lamp.

test of color vision.[16,17,19] It has also been used to test the color rendering properties of different light sources.[20–24] Each of the four series of discs is arranged in a wooden box to which two anchor discs are permanently mounted at either extreme. The two anchor discs show the starting and ending points of each series. The task of the observers is to order the discs in each series according to their hue between the two anchor discs. The test was originally designed to be administered under relatively high light levels (illuminance > 25 fc) and under daylight or a comparable electric light source,[16] such as a CIE Illuminant C type lamp.[2] There is no time limit to perform the task or is time part of the score. A test score is based on the number of transpositions, that is, the number of incorrectly placed color discs within each series.

During this study, observers performed the complete test (four color series) under diffuse daylight conditions (illuminance > 50 fc) and under each combination of SPD and light level in all three experiments. Because a different set of observers participated in each experiment, it is not possible to compare absolute performance scores across experiments. Further, given that none of the light sources used in the study was CIE Illuminant C, an absolute measure of color discrimination could not be performed, especially when warm white light sources were used. Indeed, the interactions between the source SPDs and the spectral reflectance of the discs were not assessed. Also because one of the experimental light levels (5 fc) was below the minimum recommendation to perform the test (illuminance > 25 fc), it was, again, not possible to make absolute assessments of color discrimination for this set of conditions. Therefore, the results of this study should only be taken as relative comparisons of color discrimination under each of these specific lighting conditions.

The collage consisted of six pictures containing blue jays and red cardinals. Two of the pictures also showed human skin tones and green foliage as part of the composition. A high-quality off-screen 17 by 11 inches was printed on 29 lb semimatte white stock paper at 2400 by 2400 dots/inch using a four-color (cyan, magenta, yellow, black) emulsion aggregation toner printer (DocuColor 240PS; Xerox Corporation, Stamford, CT). The print was mounted to a foam board easel, and the angle was adjusted to prevent veiling reflections.

Observers

Twelve observers participated in the first and second experiments, and 10 observers participated in the third experiment. A total of 29 observers participated in the study, 19 males and 10 females; two observers participated in both tasks of all three experiments and two observers participated in both tasks of the first and second experiments. The rest of the observers participated in one or two of the tasks of the different experiments. All had normal color vision, as tested with the Ishihara pseudoisochromatic plates' screening method,[25] and were corrected to normal visual acuity (20/20 or better). The median age for all subjects was 32 years (range: 19–62 years, standard deviation: 13.6 years).

Procedure

Observers were instructed as to the purpose of the study at the beginning of each experiment and, if needed, a few trial presentations were conducted to aid in clarifying the procedures.

During the Farnsworth–Munsell 100-Hue Test trials, observers completed the task first under natural daylight conditions from a north window (illuminance > 50 fc) and second, under each of the four light sources in each experiment at two light levels, 5 and 50 fc (Fig. 1). The order of the light level, test series, and the light sources was randomized for each subject.

On a given paired-comparison trial, observers were presented the picture collage, and it was sequentially illuminated by two different SPDs. Observers were allowed to look at the collage under the first SPD for as long as they desired before switching to the second SPD. For this protocol, once the second SPD was presented, the collage could not be viewed again under the first SPD.

In each experiment, for both light levels, 5 and 50 fc, the collage was seen by every observer three times under the four different SPDs. All the trials for one light level were completed before completing the trials for the second light level. Light level order was counterbalanced across observers. Six possible paired comparisons, each in two sequences, can be performed for four different SPDs; the order of the 12 paired comparisons was randomly presented to observers in one of two sessions per light level. For each SPD paired comparison, observers were first asked which SPD rendered the collage, overall, more ''vividly.'' They were then asked which SPD rendered the reds in the collage more ''vividly,'' and finally they were asked which SPD rendered the blues more ''vividly.'' Similarly, during a second session for the same light level, observers were asked to select the light source they thought better rendered the collage overall, as more ''natural.'' They next evaluated the reds and then the blues as more or less ''natural.'' In total, every observer performed 144 paired comparisons for each experiment (12 possible paired comparisons $\times$ 3 replications $\times$ 2 light levels [5 and 50 fc] $\times$ 2 questions [vividness and naturalness] = 144 presentations).

The experimental sessions were scheduled at the convenience of the observers and were completed over several days for each observer.

RESULTS

Farnsworth–Munsell 100-Hue Test

Figure 3 shows the results of the Farnsworth–Munsell 100-Hue Test for the three experiments in successive panel rows. Every panel in Fig. 3 shows the total score

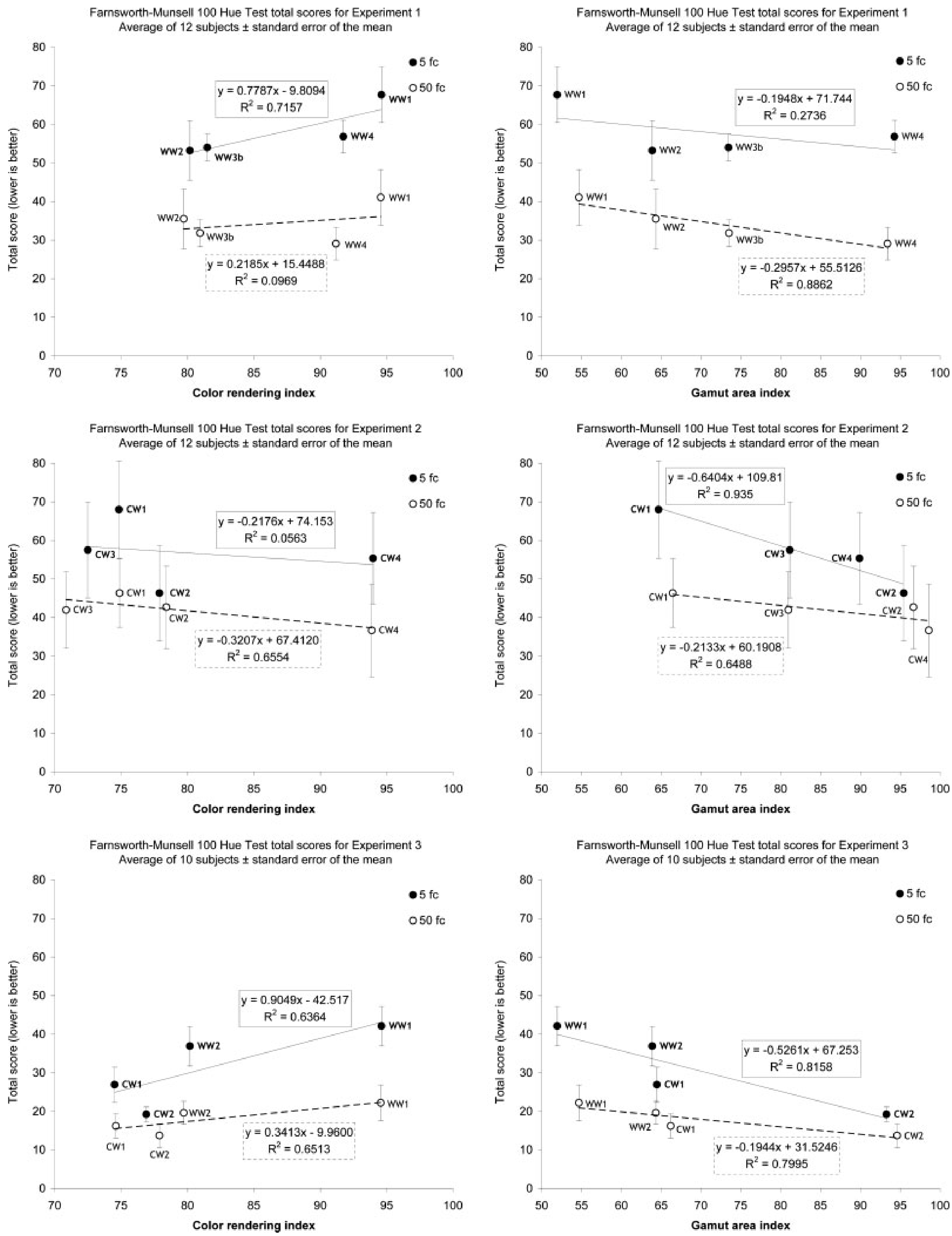

FIG. 3. Farnsworth–Munsell 100 Hue Test scores for Experiments 1–3 plotted as a function of CRI and GAI for both experimental light levels.

for both light levels, 5 and 50 fc, plotted as a function of CRI (left panels) and GAI (right panels). Lower total scores in the Farnsworth–Munsell system indicate better color discrimination.

What can be readily appreciated from this figure is that light level is important for color discrimination. Total scores are consistently lower at 50 fc than at 5 fc for all light sources in every experiment. It is also clear from this figure that GAI is better than CRI as a predictor of color discrimination. The slopes for the six linear regressions (two light levels in three experiments) relating total score and GAI are always negative, as would be expected if increasing GAI gave better color discrimination. Both positive and negative slopes were observed for CRI, suggesting that CRI is unrelated or at least inconsistently related to color discrimination. Another way of comparing the predictive power of each metric is by looking at their ability to rank order the four light sources in each experiment. A total of four transpositions of the discs (errors) are seen in the ranking of light sources with GAI, whereas 15 transpositions are observed with CRI. By using a simple scoring system based on the Farnsworth-Munsell scoring procedure, the total error scores would be 12 and 68 for GAI and CRI, respectively.

Paired Comparisons

Figure 4 shows the results of the paired comparisons for the three experiments in successive panel rows. Every panel in Fig. 4 shows the average percentage of times one light source was chosen over the others in the paired comparisons. Every panel separately shows the percentages for judgments of reds, blues, and overall, grouped by light level; data for 5 fc are presented on the left half of a panel and data for 50 fc are presented on the right. Paired comparisons using ''vividness'' as the judgment criterion are shown in the left panels of Fig. 4; those for ''naturalness'' are shown in the right panels.

The average preference percentages for different stimuli (red objects, blue objects, and overall) at different light levels (5 or 50 fc) and for different subjective criteria (''vividness'' or ''naturalness'') were then fitted as a function of CRI and of GAI to determine how the two metrics related to the paired comparison data. Linear regressions were used to evaluate the different relationships. Table II shows the results of these linear regressions. Table III shows the number of negative ($m < 0$) or ''zero'' ($|m| < 0.001$) slopes from the linear regressions of Table II and the probability that each proportion would occur by chance. It can be concluded from Table III then that GAI is a good predictor of responses when objects are illuminated by warm white sources (Experiment 1) and that CRI is a good predictor of responses under cool white sources (Experiment 2), but not vice versa.

Figure 5 illustrates another aspect of the need for more than one metric to ensure good color rendering. This figure shows a sample of the regressions from Table II illustrating the observation that CRI and GAI are differentially related to subjective judgments object colors. Specifically in Experiment 3, where mixed CCTs were used, CRI was predictive of subjective judgments of ''vividness'' for reds ($P < 0.05$), but is not predictive of subjective judgments for blues. Conversely, GAI appears to be predictive of subjective judgments of ''vividness'' for blues ($P < 0.05$), but not for reds. Moreover, in the particular case illustrated in Fig. 5, there is a *negative* correlation between subjective judgments of blues for CRI as well as between subjective judgments of reds for GAI. Chi-square statistical tests were also performed on the data in Fig. 5 to determine if the paired comparison rankings were systematically related to CRI and GAI in Experiment 3. At 50 fc when subjects used ''vividness'' as the evaluation criterion, WW1, the source with the highest value of CRI, was statistically ranked higher ($X^2 = 49.4$, $P < 0.001$) than the other three sources for red objects (upper left panel) and CW2, the source with the highest value of GAI, was statistically ranked higher ($X^2 = 21.0$, $P < 0.001$) than the other three sources for blue objects (lower right panel).

Another, interesting and important observation from Fig. 4 is the direct comparisons between warm light sources and cool light sources (Experiment 3) for different object colors (red or blue) when observers were asked to base their subjective judgments on ''vividness.'' Warm light sources were chosen more times over cool light sources for red objects. Conversely, cool light sources were chosen more times over warm light sources for blue objects. The trends for overall judgments are ambiguous, but seem to be mainly based upon the stronger judgment criterion; namely, when seen under cool light sources, overall judgments of the collage appear to be based upon the color rendering of blue objects whereas when the collage is seen under warm light sources, overall judgments appear to be based upon the color rendering of red objects. This observation is more or less true for judgments of ''naturalness,'' but the difference in average preference percentages between light sources is weaker.

DISCUSSION

CRI is Simply not Enough

Sixty years ago Bouma, in his description of the color rendering properties of daylight, articulated the color rendering characteristics important for any light source. A light source with good color rendering properties should support a full range of object color perceptions, should enable good color discrimination, and should not make colors look unnatural. Fifty years ago, CRI was developed to address the last of these characteristics. Using familiar light sources as reference light sources (incandescent and daylight), the developers of CRI believed that electric light sources that minimized shifts in the chromaticities of object colors would minimize color distortions and thereby make object colors look more ''natural.'' CRI was clearly acknowledged by its authors as a single metric purportedly meeting only one aspect of color rendering

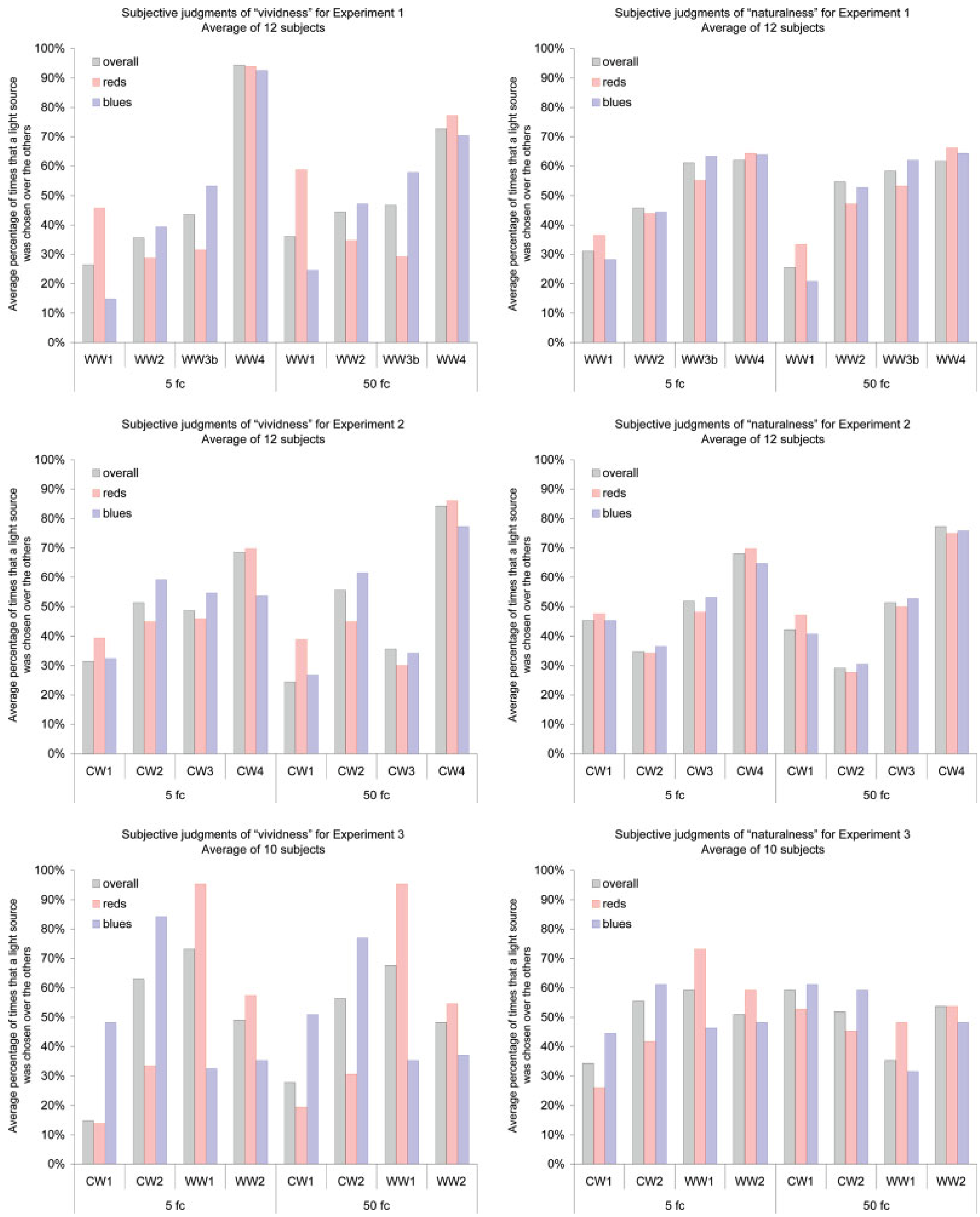

FIG. 4. Percentage of times that a light source was chosen over the others during the paired comparisons. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

(''naturalness'') and therefore was acknowledged as an incomplete measure of the color rendering properties of a light source. Some decades later, however, CRI is presently the main criterion used by the lighting industry to describe the color rendering properties of electric light sources and it is the only one widely accepted by practitioners as such.[10] Certainly the data presented here and the arguments presented over the last 50 years, support

TABLE II. Linear ($y = mx + b$) regressions relating the paired comparison rankings (for red and blue objects and for the overall assessment) to CRI and to GAI, at each light level (5 and 50 fc) and in each Experiment (1, 2, and 3).

	Naturalness 5 fc			Naturalness 50 fc			Vividness 5 fc			Vividness 50 fc		
	m	b	R^2	m	b	R^2	m	b	R^2	m	b	R^2
Experiment 1 WW1, WW2, WW3b, WW4												
CRI												
Overall	−0.0076	1.1655	0.1425	−0.0135	1.6706	0.3618	0.0104	−0.4085	0.0617	0.0030	0.2361	0.0203
Red	−0.0014	0.6177	0.0064	−0.0035	0.8050	0.0364	0.0269	−1.8438	0.4156	0.0255	−1.7112	0.7160
Blue	−0.0086	1.2521	0.1340	−0.0155	1.8414	0.3249	−0.0007	0.5632	0.0003	−0.0082	1.2128	0.0979
GAI												
Overall	0.0073	−0.0150	0.7858	0.0079	−0.0678	0.6243	0.0163	−0.6580	0.9229	0.0093	−0.1674	0.9522
Red	0.0067	0.0244	0.9655	0.0081	−0.0781	0.9582	0.0126	−0.3918	0.5538	0.0063	0.0475	0.2199
Blue	0.0084	−0.0973	0.7780	0.0099	−0.2071	0.6620	0.0182	−0.7868	0.9978	0.0111	−0.2953	0.8939
Experiment 2 CW1, CW2, CW3, CW4												
CRI												
overall	0.0104	−0.3316	0.5252	0.0145	−0.6517	0.5122	0.0128	−0.5218	0.6716	0.0240	−1.4052	0.8469
red	0.0119	−0.4523	0.6163	0.0131	−0.5385	0.4599	0.0133	−0.5594	0.9006	0.0246	−1.4580	0.9956
blue	0.0083	−0.1632	0.4522	0.0136	−0.5808	0.4906	0.0031	0.2509	0.0640	0.0205	−1.1300	0.7671
GAI												
overall	0.0003	0.4753	0.0008	0.0040	0.1609	0.0856	0.0090	−0.2446	0.6354	0.0156	−0.8377	0.8044
red	−0.0002	0.5189	0.0004	0.0019	0.3331	0.0229	0.0051	0.0766	0.2563	0.0106	−0.4054	0.4101
blue	0.0003	0.4783	0.0009	0.0042	0.1389	0.1055	0.0084	−0.1948	0.8867	0.0147	−0.7600	0.8833
Experiment 3 CW1, CW2, WW1, WW2												
CRI												
Overall	0.0085	−0.1881	0.4621	−0.0122	1.4867	0.9806	0.0205	−1.1620	0.5069	0.0162	−0.8173	0.6638
Red	0.0203	−1.1438	0.7581	−0.0017	0.6375	0.1326	0.0376	−2.5519	0.8986	*0.0380*	*−2.5798*	*0.9080*
Blue	−0.0023	0.6876	0.0732	−0.0153	1.7399	0.8985	−0.0140	1.6385	0.2708	*−0.0113*	*1.4140*	*0.2445*
GAI												
Overall	0.0004	0.4759	0.0034	0.0022	0.3462	0.1446	0.0007	0.4500	0.0028	−0.0002	0.5109	0.0003
Red	−0.0059	0.8982	0.2686	−0.0014	0.5943	0.3786	−0.0113	1.2597	0.3363	*−0.0118*	*1.3179*	*0.3887*
Blue	0.0038	0.2415	0.8392	0.0052	0.1387	0.4630	0.0130	−0.3761	0.9686	*0.0107*	*−0.2377*	*0.9666*

Cells in *italic* face correspond to the conditions plotted in Figure 5.

the arguments made by its developers, namely, CRI is not a universal measure of color rendering.

Reflecting these early arguments and anticipating the data presented here, Rea and coworkers[10] put forward the notion that three metrics could be used for assessing the color rendering properties of light sources: CRI, GAI, and FSCI. That document suggested that each of the three metrics might be used to characterize different aspects of color rendering, naturalness (CRI), saturation (GAI), and discrimination (FSCI).[†] The data presented here, however, indicate for the first time that a simple homology between a given color rendering metric (e.g., GAI) and a color rendering criterion (e.g., ''vividness'') is not perfect. Rather, it appears that the metrics must be used together to ensure good color rendering. The two sources with the highest values of CRI as well as GAI were always selected over the other sources in Experiments 1 and 2. In Experiment 1, WW4 was chosen every time (12 out of 12 times; $p < 0.00024$) when it was compared with other sources for both light levels (5 and 50 fc), all three stimuli (red objects, blue objects, and overall), and for both evaluation criteria (''naturalness'' and ''vividness''). CW4 was chosen in 11 of the 12 comparisons in Experiment 2; the only time that it was not chosen was for the ''vividness'' judgment criterion of blue objects at 5 fc, but importantly, under that condition CW2 actually had a higher value of GAI than CW4 and, as previously discussed, GAI is a better predictor of color vividness than CRI for blue objects. Thus, in nearly every case (23 out of 24 times) the preferred light source in the paired comparisons had the highest values of both GAI and CRI.

It is also worth noting, although the results are not shown here, that among the individual CRI submetrics, R_9, which is purported to characterize the color rendering capabilities of a light source for red objects was not more predictive of the paired comparison data for the ''vividness'' of red objects than was the overall value of CRI itself. Moreover, R_{12}, purported to characterize the color rendering capabilities of a light source for blue objects, was not at all predictive of the paired comparison data for the ''vividness'' of blue objects. Thus, it appears that GAI is better at predicting subjective judgments for blue objects than CRI or any of its submetrics.

In terms of color discrimination, the data in Fig. 3 demonstrate that CRI is not useful at predicting performance under the conditions tested. Rather, it appears that GAI is always more predictive of color discrimination

[†]The light sources used in the present study gave highly correlated vales of GAI and FSCI, so it was not possible to assess independently the utility of GAI and FSCI as measures of color rendering. It was clear from the present data, however, that at least two metrics, CRI and either GAI or FSCI should be used to represent the color rendering properties of broad-band, white light sources used as light sources. GAI was featured as the adjunct metric to CRI over FSCI in this report because of the historical priority given to gamut area over the recently developed FSCI and because the same reference color chips are used to calculate both CRI and GAI.

TABLE III. Proportions of negative (opposite expected trend; $m < 0$) and zero (no trend; $|m| < 0.001$) slopes from the linear regressions in Table II.

	CRI	GAI
Experiment 1 WW1, WW2, WW3b, WW4	8 out of 12 (from 0 to 8 inclusive, $P = 0.927$)	*0 out of 12* ($P = 0.00024$)
Experiment 2 CW1, CW2, CW3, CW4	*0 out of 12* ($P = 0.00024$)	3 out of 12 (from 0 to 3 inclusive, $P = 0.073$)
Experiment 3 CW1, CW2, WW1, WW2	6 out of 12 (from 0 to 6 inclusive, $P = 0.613$)	7 out of 12 (from 0 to 7 inclusive, $P = 0.806$)

The binomial probabilities that those proportions reflect trends in the expected direction ($m > 0$) by chance alone are in brackets adjacent to each proportion. Assuming a conventional chance probability (P) = 0.05, only the two proportions in *italic* face would not have occurred by chance, indicating that there is a statistically reliable relationship between the color rendering metric (CRI in Experiment 2 and GAI in Experiment 1) and subjective preferences.

performance than CRI. Nevertheless, the light sources that had high values of CRI as well as GAI (WW4 and CW4) were superior for color discrimination as well as for the paired comparisons.

The central thesis of this article then is that CRI alone will not always satisfactorily represent the color rendering properties of a light source and that another color rendering metric, GAI, should be used in conjunction with CRI in the specifications for illuminating spaces and objects where color rendering is important. The data presented here are consistent with current lighting recommendations for illuminating neonatal intensive care units (NICUs),

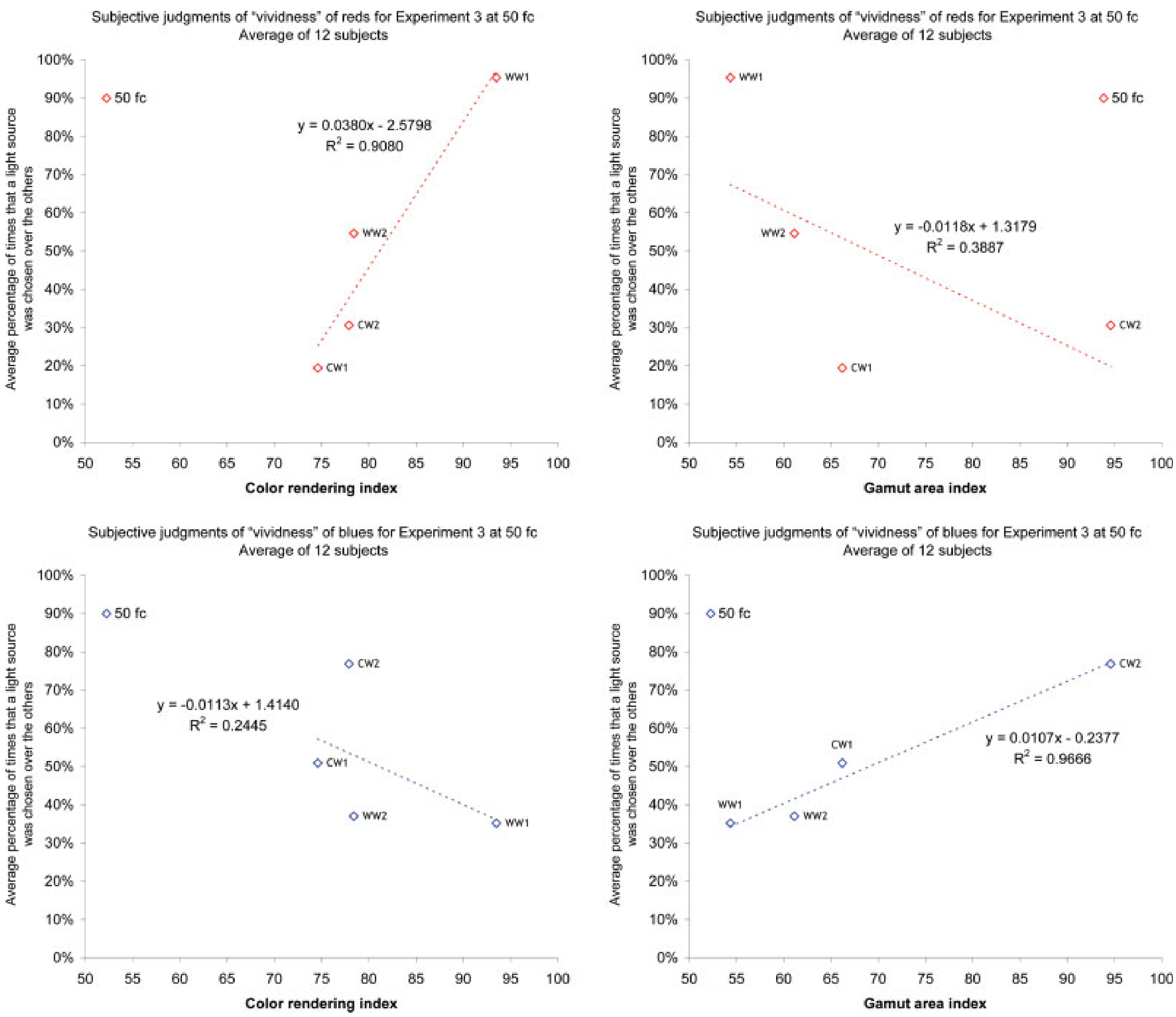

FIG. 5. Relationships between CRI (left panels) or GAI (right panels) and the average percentage of times that a light source was chosen over the others using subjective judgments of "vividness" for reds and blues at 50 fc. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

namely, lamps should have a minimum CRI value of 80, a minimum FSCI of 55, and a GAI between 80 and 100.[26‡] Sources meeting the CRI, the GAI and the FSCI criteria proposed in the NICU lighting standard (i.e., WW4 and CW4) appear to always provide superior color discrimination and higher ratings of ''vividness'' and ''naturalness.'' It remains to be determined, however, if FSCI provides any utility beyond CRI and GAI for characterizing the color rendering properties of electric light sources. The interactions between CRI and GAI also remain to be studied. For example, it remains to be determined whether a high (and how high) value of GAI can compensate for a low value of CRI, or vice versa.

Additional Insights

Irradiance level is important for good color rendering. In addition to meeting the NICU color rendering criteria, it is necessary to have a sufficient illuminance level to ensure good color rendering. The literature consistently shows that illuminance levels of 100 fc or more are needed for the best performance on the Farnsworth–Munsell 100 hue test.[20,22,24] This study showed that color discrimination was always better at 50 fc than at 5 fc. Based upon the literature, even higher levels are probably needed for still better color discrimination.

The two subjective color rendering criteria used by observers in this study, ''naturalness'' and ''vividness,'' do not appear to be remarkably different in terms of their relationships to the two color rendering metrics, CRI and GAI. As shown in Fig. 4, the paired comparison data are similar for both subjective criteria. It does appear, however, that ''naturalness'' is a more ambiguous subjective criterion than ''vividness'' for observers to understand and respond to in the experiment as reflected in the less distinct differences among the various light sources in the paired comparison data (Fig. 4). This inference was supported by informal reports by subjects at the conclusion of the experiment. They often commented that they did not know exactly what was meant by ''naturalness.'' Therefore, ''naturalness,'' despite its importance in the development of CRI, may not be a particularly meaningful subjective criterion for color rendering under broadband ''white'' sources.

CCT is also apparently important to color rendering in at least two ways. First, red objects appear to be more important for subjective assessments than blue objects when they are illuminated by a warm CCT; the opposite is true when blue and red objects are illuminated by a cool CCT. These data indirectly reinforce the idea that the CCT of the lamp should ''complement'' object colors. This is no surprise to professional interior designers; warm light sources should be used with red and yellow fabrics and paints whereas cool light sources should be used with blue and green fabrics and paints. Second, when it is nevertheless important to render blue objects well under a warm light source, GAI but not CRI, will be predictive of their perceived ''vividness'' and ''naturalness.'' The opposite is true for rendering red objects under a cool light source; CRI, but not GAI, will be predictive of ''vividness'' and ''naturalness.'' Because most scenes have many object colors, it is important that high values of both CRI and GAI be used to ensure that the full range of object colors appear vivid and natural, and to provide good color discrimination.

The recent and rapid development of light emitting diodes (LEDs) has sparked new interest in color rendering.[10,14,15] Several studies have pointed out the limitations of CRI when narrow-band light sources, such as red-green-blue LEDs are used to produce white light.[12,15] Although LEDs were used in this study, a mixture of narrow-band light sources might lead to different results than the ones presented here, but it seems highly unlikely that sources meeting both the CRI and GAI criterion values recommended for NICUs would render object colors poorly. Nevertheless, this assertion should be explicitly tested.

Finally, the concept of good color rendering is undoubtedly situational. Those who developed CRI were very interested in making objects appear ''natural.'' Judd noted that ''flattery'' could be another important color rendering criterion. His ''flattery index'' was never adopted, but it is clear that enhanced, or ''unnatural,'' colors can also be important in some applications. One practical example illustrates this point. It is often important for a butcher to ''enhance'' the redness of hamburger in a meat case, even though it appears ''unnaturally'' red in this situation. Clearly, however, a ''meat lamp'' would be entirely inappropriate in the NICU. Therefore, the concept of good color rendering is dependent upon the application and not a rigid, immutable characteristic of the spectral power distribution of a light source. To reinforce the hopes articulated by the developers of CRI, more sophisticated attempts should be undertaken to understand how to control SPD for a variety of intended purposes.[27,28]

SUMMARY

Although accepting the two-metric proposal made here will not necessarily be a far, far better thing than the lighting industry has ever done before, it would appear to be a useful improvement over reliance on CRI as the only measure of color rendering. Certainly a number of authors have made the point that CRI is insufficient as the sole measure of color rendering for electric light

[‡]GAI has been normalized to an equal energy spectrum value of 100, but values higher than this have been shown to ''over emphasize'' or ''distort'' certain hues. For example the Promolux lamp utilized in this study is used to enhance the redness of meat in a butcher's display case. The color ''distortion'' from this lamp contradicts one of the original tenets of color rendering, namely that a light source should make objects appear ''natural.'' Notwithstanding the apparent problems with this subjective criterion being used as a basis for color rendering, as discussed earlier, GAI values greater than 100 might make meat or, more importantly, premature infants, appear ''unnatural.'' Therefore, a GAI upper limit of 100 was recommended in the NICU standard and, absent additional experimental data, recommended here as well.

sources.[3,4,9,10,14,27–29] And it should be recalled that the authors of CRI made this same point over 50 years ago. Still, after half a century, there is no other metric generally accepted or officially promulgated as a replacement for or a compliment to CRI. The present study suggests that gamut area, or more precisely GAI, can be a useful, practical adjunct to the well-established CRI in ensuring both good color discrimination and satisfactory perceptions of object color ''vividness'' and ''naturalness.'' These data also reinforce the color rendering recommendations for NICUs[26]; a light source providing a minimum CRI of 80 and a GAI between 80 and 100 will provide good color discrimination and make objects in the scene appear both ''vivid'' and ''natural,'' given sufficient illuminance is provided to the visual scene. Finally, it should be stressed that although this two-metric proposal appears to be both useful and practical, additional studies should be conducted to determine if these two metrics are ideal and are sufficient for ensuring good color rendering for architectural lighting in general and for niche applications in particular.

ACKNOWLEDGMENTS

The authors are gratefully acknowledge Yukio Akashi, Andrew Bierman, Mariana Figueiro, and Russ Leslie for providing helpful comments; Robert Wolsey and Rui Qi are also gratefully acknowledged for their help in conducting part of the experiments as is John Bullough for providing valuable insight during the planning of the experiment and for his help with the statistical analysis.

APPENDIX

Calculation of Gamut Area Index

Gamut area of a light source is commonly calculated as the area of the polygon defined by the chromaticities in CIE 1964 color space of the eight CIE standard color samples specified in CIE Technical Report No. 13.3-1995[2] when illuminated by a test light source. For this study, the gamut area of the equal energy spectrum is scaled to 100 and defined as gamut area index (GAI). Different light sources are scaled accordingly.[10] GAI is a convenient metric to supplement CRI because, like CRI, it is derived from the spectral power distribution of a light source and the resulting chromaticities of the same eight CIE standard color samples.[2]

1. Nickerson D. Light sources and color rendering. J Opt Soc Am 1960;50:57–69.
2. Commission Internationale de l'Éclairage (CIE). CIE publ. no. 13.3:1995. Method of measuring and specifying colour rendering properties of light sources. Vienna, Austria: CIE Central Bureau; 1995. 16 p. REF CIE 13.3-1995.
3. Judd DB. A flattery index for artificial illuminants. Illum Eng 1967;62:593–598.
4. Narendran N, Deng L. Color rendering properties of LED light sources. In Ferguson IT, Narendran N, DenBaars SP, Park Y-S, editors, the Proceedings of SPIE, Volume 4776 Solid State Lighting II; Seattle, WA, 2002. p 61–67.
5. Pracejus WG. Preliminary report on a new approach to acceptance studies. Illum Eng 1967;62:663–673.
6. Thornton WA. Color-discrimination index. J Opt Soc Am 1972;62:191–194. (Available online at http://www.opticsinfobase.org/abstract.cfm?URI=josa-62-2-191).
7. Thornton WA, Chen E. What is visual clarity? J Illum Eng Soc 1978;7:85–94.
8. Boyce PR, Simons RH. Hue discrimination and light sources. Light Res Tech 1976;9:11–24.
9. Guo X, Houser KW. A review of colour rendering indices and their application to commercial light sources. Light Res Tech 2004; 36:183–199.
10. Rea M, Deng L, Wolsey R. NLPIP Lighting Answers: Light Sources and Color. Troy, NY: Rensselaer Polytechnic Institute; National Lighting Product Information Program; 2004.
11. Thornton WA. Lamps for assessing metamerism. J Illum Eng Soc 1974;3:11–18.
12. Thornton WA. The high visual efficiency of prime color lamps. Light Design Appl 1975;5:35–41.
13. Thornton WA, Corth R, Evans GS. Fluorescent light sources. Light Design Appl 1975;5:6–14.
14. Davis W, Ohno Y. Toward an improved color rendering metric. In Ferguson IT, Carrano JC, Taguchi T, Ashdown IE, Editors, the Proceedings of SPIE, Volume 5941 Fifth International Conference on Solid State Lighting; San Diego, CA, 2005. p 59411G-1-8.
15. Commission Internationale de l'Éclairage (CIE). CIE publ. no. 177:2007. Colour rendering of white LED light sources. Vienna, Austria: CIE Central Bureau; 2007. 14 p.
16. Farnsworth D. The Farnsworth-Munsell 100-Hue and dichotomous tests for color vision. J Opt Soc Am 1943;33:568–578.
17. Farnsworth D. The Farnsworth-Munsell 100-Hue test for the Examination of Colour Discrimination: Manual. Baltimore, MD: Munsell Color Company; 1957.
18. Rea M, Deng L, Wolsey R. NLPIP Lighting Answers: Full-Spectrum Light Sources. Troy, NY: Rensselaer Polytechnic Institute; National Lighting Product Information Program; 2003 (revised March 2005).
19. Viliūnas V, Lukauskiene R, Švegžda A, Žukauskas A. End-box scoring artefact evaluation of the Farnsworth-Munsell 100-Hue colour vision test. Ophthalmic Physiol Optic 2006;26:580–586.
20. Boyce PR. Illuminance, lamp type and performance on a colour discrimination task. Light Res Tech 1976;8:195–199.
21. Donaldson GB. Instrumentation for the Farnsworth-Munsell 100-hue Test. J Opt Soc Am 1977;67:248–249.
22. Bowman KJ, Cole BL. A recommendation for illumination of the Farnsworth-Munsell 100-hue test. Am J Opt Phys Opt 1980;57:839–843.
23. Verriest G, Van Laethem J, Uvijls A. A new assessment of the normal ranges of the Farnsworth-Munsell 100-hue test scores. Am J Ophthamol 1982;93:635–642.
24. Knoblauch K, Saunders F, Kusuda M, Hynes R, Podgor M, Higgins KE, de Monasterio FM. Age and illuminance effects in the Farnsworth-Munsell 100-hue test. Appl Opt 1987;26:1441–1448.
25. Ishihara, S. Tests for Colour-Blindness. Handaya, Tokyo: Hongo Harukicho; 1917.
26. Figueiro MG, Appleman K, Bullough JD, Rea MS. A discussion of recommended standards for lighting in the NICU. J Perinatol 2006;26:S19–S26.
27. Worthey JA. Color rendering: Asking the question. Color Res App 2003;28:403–412.
28. Worthey JA. Color rendering: A calculation that estimates colorimetric shifts. Color Res App 2004;29:43–56.
29. van Trigt C. Color rendering, a reassessment. Color Res App 1999;24:197–206.